Characterization of genetic variation and yield heterosis in *Cucumis melo*

Gur, A. Agricultural Research Organization

Buckler, E.S. Cornell University

Burger, J. Agricultural Research Organization

Tadmor, Y. Agricultural Research Organization

Klapp, I. Agricultural Research Organization

Project award year: 2016
Three year research project

Abstract

Project objectives: **1)** Characterization of variation for yield heterosis in melon using Half-Diallele (HDA) design. **2)** Development and implementation of image-based yield phenotyping in melon. **3)** Characterization of genetic, epigenetic and transcriptional variation across 25 founder lines and selected hybrids. The epigentic part of this objective was modified during the course of the project: instead of characterization of chromatin structure in a single melon line through genome-wide mapping of nucleosomes using MNase-seq approach, we took advantage of rapid advancements in single-molecule sequencing and shifted the focus to *Nanopore* long-read sequencing of all 25 founder lines. This analysis provides invaluable information on genome-wide structural variation across our diversity **4)** Integrated analyses and development of prediction models

Agricultural heterosis relates to hybrids that outperform their inbred parents for yield. First generation (F1) hybrids are produced in many crop species and it is estimated that heterosis increases yield by 15-30% globally. Melon (*Cucumis melo*) is an economically important species of The Cucurbitaceae family and is among the most important fleshy fruits for fresh consumption Worldwide. The major goal of this project was to explore the patterns and magnitude of yield heterosis in melon and link it to whole genome sequence variation. A core subset of 25 diverse lines was selected from the Newe-Yaar melon diversity panel for whole-genome re-sequencing (WGS) and test-crosses, to produce structured half-diallele design of 300 F1 hybrids (*MelHDA25*). Yield variation was measured in replicated yield trials at the whole-plant and at the rootstock levels (through a common-scion grafted experiments), across the F1s and parental lines. As part of this project we also developed an algorithmic pipeline for detection and yield estimation of melons from aerial-images, towards future implementation of such high throughput, cost-effective method for remote yield evaluation in open-field melons. We found extensive, highly heritable root-derived yield variation across the diallele population that was characterized by prominent best-parent heterosis (BPH), where hybrids rootstocks outperformed their parents by 38% and 56 % under optimal irrigation and drought-stress, respectively. Through integration of the genotypic data (~4,000,000 SNPs) and yield analyses we show that root-derived hybrids yield is independent of parental genetic distance. However, we mapped novel root-derived yield QTLs through genome-wide association (GWA) analysis and a multi-QTLs model explained more than 45% of the hybrids yield variation, providing a potential route for marker-assisted hybrid rootstock breeding. Four selected hybrid rootstocks are further studied under multiple scion varieties and their validated positive effect on yield performance is now leading to ongoing evaluation of their commercial potential. On the genomic level, this project resulted in 3 layers of data: 1) whole-genome short-read Illumina sequencing (30X) of the 25 founder lines provided us with 25 genome alignments and high-density melon HapMap that is already shown to be an effective resource for QTL annotation and candidate gene analysis in melon. 2) fast advancements in long-read single-molecule sequencing allowed us to shift focus towards this technology and generate ~50X *Nanopore* sequencing of the 25 founders which in combination with the short-read data now enable *de novo* assembly of the 25 genomes that will soon lead to construction of the first melon pan-genome. 3) Transcriptomic (3' RNA-Seq) analysis of several selected hybrids and their parents provide preliminary information on differentially expressed genes that can be further used to explain the root-derived yield variation. Taken together, this project expanded our view on yield heterosis in melon with novel specific insights on root-derived yield heterosis. To our knowledge, thus far this is the largest systematic genetic analysis of rootstock effects on yield heterosis in cucurbits or any other crop plant, and our results are now translated into potential breeding applications. The genomic resources that were developed as part of this project are putting melon in the forefront of genomic research and will continue to be useful tool for the cucurbits community in years to come.

Summary Sheet

Publication Summary

PubType	IS only	Joint	US only
Reviewed	1	1	0

Training Summary

Trainee Type	Last Name	First Name	Institution	Country
Ph.D. Student	Dafna	Asaf	Hebrew University of Jerusalem	Israel
Ph.D. Student	Oren	Elad	Hebrew University of Jerusalem	Israel

Collaboration:

During the course of this research, the Israeli teams has focused on the melon diversity, populations development, field experiments, aerial-image analyses and genetic analyses. The US team worked on aspects of sequence and genomic analyses and bioinformatics pipelines. The project clearly benefitted from this complementary division of the project objectives. The ability to generate, in parallel, large-scale phenotypic and genomic data on diverse melon population, is the main strength of this project. Periodic updates conversations took place between the teams and PI and Co-PI. These includes ongoing assessments of the results and right channels of actions. An important milestone in the collaboration during the second year was a one-month visit of Elad Oren, a PhD student from the Gur Lab at ARO, to the Buckler Lab at Cornell. This visit was important for analysis of the project data and at the same time was very significant for importing knowhow and bioinformatics tools to be implemented by the Gur lab. The collaboration has not ended yet, as we expect three additional joint publications to be submitted based on the results.

Abstracts from this collaborative project were submitted to the following meetings:

- A Kalantar, A Dashuta, Y Edan, A Dafna, A Gur, I Klapp. Estimating melon yield for breeding processes by machine-vision processing of UAV images. European Confference of Precision Agriculture (ECPA) 2019. Montpellier, France. Oral Presentation.
- Rees Evan, Oren Elad, Song Baoxing, Wang Xiaoyun, Tate Ann, Glaubitz Jeffrey, Romay Cinta and Buckler Edward. A fast computational pipeline for de novo assembly of plant genomes from long reads. Maize Genetics Meeting 2019. Poster#5. March 2019, St. Louis, USA.
- Elad Oren, Galil Tzuri, Evan R. Rees, Baoxing Song, Arthur Schaffer, Yaakov Tadmor, Joseph Burger, Edward Buckler and Amit Gur. Panning the Melon Genome. Submitted to Eucarpia Cucurbitaceae 2020. Accepted for oral presentation. Conference planned to be held on May 2020 in Almeria, Spain and postponed due to COVID-19.
- Asaf Dafna, Elad Oren, Ilan Halperin, Galil Tzuri, Ayala Meir, Tal Isaacson, Naomi Ori, Edward Buckler, Amit Gur. "Six Feet under": Genetics and Breeding for yield-enhancing rootstocks in melon. Submitted to annual conference of the Israeli Society of crop and vegetable science. December 2020. Zoom Meeting.

1

Achievements

- *Resequencing of 25 diverse melon lines and Construction of HapMap*: Twenty-five inbred lines representing the two melon subspecies and 11 horticultural groups, were selected as the core subset for this project as described by Gur *et al.* (2017). Deep (~30X) whole-genome resequencing and alignments to the reference melon genome allowed us to build a dense haplotype map (HapMap) composed of ~4,000,000 SNPs and short InDels, that essentially cover most of the reference-based variation across our 25 founders. While this resource is fundamental to the root-derived yield GWA analysis (described below), it is also an important permanent tool for genomic analyses in current and future research in our lab and other melon research groups. This is well demonstrated in our recent publication, where we used the HapMap for detailed genomic analysis of candidate gene and causative SNPs associated with pigment accumulation in melon (Oren et al. 2019).

- *Creation of diverse, 25-way, diallele population:* the center of this project is a multi-allelic population of 300 F1 hybrids, built through a half-diallele crossing scheme between the 25 diverse founders (**Figure 1**). In addition to the effectiveness of this design for studying heterosis, it is also demonstrated as a valuable resource for broad characterization of mode-of-inheritance of fruit quality traits in melon. Additionally, this population can be used for allelism testing of recessive traits as we did in the characterization of the light rind trait (Oren et al. 2019). Another advantage of this population is the whole-genome sequence that is inferred for all 300 F1 hybrids based on their parental genomes, and allow performance of high resolution GWAS. Taken together, combined with the genome-wide parental HapMap, the diallele population is a significant product of this project that serve as pivotal resource in our current and future research.

- *Mode of inheritance of root-derived ("underground") and whole-plant yield in melon*: During the 1st year, a core subset of the diallele population was tested in replicated yield trials where each entry was tested as non-grafted plants and as rootstocks grafted with common scion. This replicated yield trial of more than 3,000 plants exposed prominent best-parent yield heterosis on both non-grafted and grafted (rootstock) experiments that were not correlated with each other, indicating on the involvement of multiple avenues for yield heterosis, including important role for the roots **(Figure 2)**. Additionally, we could identify in this experiment F1 hybrid rootstocks that improved yield by 25% compared to an elite commercial hybrid. The significant and heritable root-derived yield heterosis on this experiment promoted the focus on rootstock heterosis in the following years.

- *Development of an algorithmic pipeline for detection and yield estimation of melons from aerial-images, towards future implementation of high throughput, cost-effective method for remote yield*

evaluation in open-field melons: Through the collaboration in this project with Dr. Iftach Klapp from the institute of agricultural engineering at ARO, a preliminary pipeline was developed based on aerial-images and data collected on the field trials. The yield estimation provides both the number of melons and the weight of each melon. The system includes three main stages: melon detection, geometric feature extraction, and individual melon yield estimation. The melon-detection process was based on the RetinaNet deep convolutional neural network. Transfer learning was used for the training to detect small objects in high-resolution images successfully. Overall, yield estimation derived by summing the weights of all melons in the field resulted in only a 3% underestimation of the actual total yield. The algorithmic pipeline performance and development process are described in details in the recent paper by Kalantar et al. (2020).

- *Large-scale genetic analysis of "underground" heterosis in melon*: Based on the positive results obtained on the first rootstock experiment, we extended the genetic analysis and tested the complete *HDA20* set (190 half-diallele F1 hybrids + 20 parents) as rootstocks in replicated yield trial in the open-field at Newe-Ya'ar under irrigated and non-irrigated conditions (**Figure 3a,b**). The correlation between the non-irrigated and irrigated trials was high (R=0.71, **Figure 3c**) and supported the good heritability of the root-derived yield variation (H^2=0.48). This experiment validated and extended the 1st year results and provided support for prominent root-derived yield heterosis, as the average hybrid yield was 38% and 56% higher than the inbred yield across this set in the irrigated and non-irrigated fields, respectively (**Figure 4a,b**). In this research, we use the stringent genetic definition of heterosis, that is the deviation of the hybrid from the high-parent (best-parent heterosis, BPH) which is also the relevant definition from the agricultural standpoint. **Figure 5a** shows the yield of the 190 *HDA20* hybrids partitioned to best-parent (BP) and heterotic (BPH) components. Hybrids are in ascending order by their yield, and it was unexpected to find out how prevalent was root-derived yield BPH, with 130 out of the 190 hybrids showing a certain level of positive over-dominance, and 79 hybrids significantly outperforming their best-parent and displaying significant BPH (at P<0.05). It is apparent that level of (over)dominant deviation, a non-additive genetic component, is the major contributor to root-derived hybrid yield (**Figure 5b**). The average BPH across all 190 hybrids is 26%, explaining the significant overall advantage of hybrids over their BPs (**Figure 6**). The best hybrid rootstock outperformed its best-parent by 80% and improved yield compared to the self-grafted commercial hybrid by 65% under both irrigated and drought conditions. These results demonstrate the fundamentally different mode of inheritance of root-derived yield compared to commonly studied additive traits (such as TSS or fruit size, for

2

example) and emphasize the importance of pursuing this direction as a promising novel channel for yield improvement in melon.

- *GWAS analysis and identification of root-derived yield QTLs:* while we discovered that heterosis is a key genetic factor explaining significant portion of the root-derived yield variation across our *HDA20* population, the challenge remains in predicting the level of heterosis in hybrids for effective rootstock breeding. We initially showed that parental genetic distance is a poor predictor and to further utilize the phenotypic and genotypic data collected on the *HDA20* population for development of prediction tools, we performed GWA analysis. **Figure 7** is a Manhattan plot showing multiple SNP markers across the genome that are significantly associated with hybrids root-derived yield variation. Through multivariate analysis we were able to build a model with 9 significant markers that jointly explained 48% of the root-derived yield variation across *HDA20* population. **Figure 8** is plotting the combined haplotypic effect of two significant markers that jointly explain 38% of the variation and selection for heterozygotes in both markers is associated with 15% root-derived yield improvement compared to the population mean. This is an important proof-of-concept for genetic marker based prediction of hybrid rootstock yield performance.

- *Selected hybrid rootstocks - potential product for yield improvement in melon:* based on the large scale genetic analysis of rootstocks performance under two environments, we were able to select four high-yielding hybrid rootstocks for further testing under four different scion varieties. **Figure 9** is showing the results from field experiments on 2019 and the overall advantage of these rootstocks over the different non-grafted varieties, and highlight the commercial potential of the germplasm and results that were obtained in this project.

- *Nanopore sequencing and de novo assembly of 25 melon genomes:* to address another genomic layer in our diverse set of 25 founder lines, we are performing at Cornell deep (~70X) long-read sequencing using the *Nanopore* single-molecule technology. We are then implementing a bioinformatic pipeline that integrate the Ilumina short-read sequences with the *Nanopore* long-reads to perform *de novo* assembly of the 25 genomes. While we are currently sequencing the final batch of the 25 lines to achieve ~70X depth, *de novo* assemblies of 24 genomes and comparative analysis using the currently available ~50X data is already revealing interesting results. **Figure 10** is presenting the variation in genomes size that is associated (as expected) with genetic relatedness, and independent from technical aspect of sequencing coverage. **Figure 11** and **Figure 12** are showing specific comparisons between two of the founder lines and structural variation between them.

3

Changes to research plan:

The main change to the initial project objectives and work plan relates to objective #3.4 (Characterization of chromatin structure through MNase-seq analysis on root and shoot of reference genome). This objective was modified during the course of the project: instead of characterization of chromatin structure in a single melon line through genome-wide mapping of nucleosomes, using MNase-seq approach, we took advantage of the recent rapid advancements in single-molecule sequencing and shifted the focus to *Nanopore* long-read genomic sequencing of all 25 founder lines. A combination of the *Ilumina* short reads with the *Nanopore* long reads allow *de novo* assembly of the 25 founder genomes. Comparative analysis of these 25 *de novo* assembled genomes can explore structural genomic variation across our diverse core panel, towards the creation of the first melon pan-genome. This analysis fits well with the scope of our project and allow us to test the association of such genomic variation with expression of heterosis across our diallele population.

1

Publications for Project IS-4911-16

Status	Type	Authors	Title	Journal	Vol:pg Year	Coun
Published	Reviewed	*Elad Oren, Galil Tzuri, Lea Vexler, Asaf Dafna, Ayala Meir, Adi Faigenboim, Merav Kenigswald, Vitaly Portnoy, Arthur A Schaffer, Amnon Levi, Edward S Buckler, Nurit Katzir, Joseph Burger, Yaakov Tadmor, Amit Gur*	The multi-allelic APRR2 gene is associated with fruit pigment accumulation in melon and watermelon	***Journal of Experimental Botany***	70 : 3781-3794 2019	Joint
Published	Reviewed	*Aharon Kalantar, Yael Edan, Amit Gur, Iftach Klapp*	A deep learning system for single and overall weight estimation of melons using unmanned aerial vehicle images	***Computers and Electronics in Agriculture***	178 : 2020	IS only

BARD IS-4911-16 Final Report - Appendix

Content:

Table 1: Properties of 25 *de novo* genome assemblies.

Figure 1: Our diversity exploitation path: development of core panel and half-diallele population.

Figure 2: Mode of inheritance of non-grafted and rootstocks (grafted) yield across 45 Half-diallele F1 hybrids and their parental lines – Summer 2017.

Figure 3: *HDA20* non-irrigated and irrigated yield trials in summer 2018.

Figure 4: *HDA20* grafted rootstock yield trial. Comparison of hybrids and parents in the irrigated and non-irrigated fields.

Figure 5: Partition of hybrid yield to parental and heterotic components.

Figure 6: Genetic components of hybrid yield in the *HDA20* grafted rootstock experiment.

Figure 7: GWAS of root-derived yield across 190 *HDA20* hybrids.

Figure 8: Combined effect of two QTLs associated with root-derived yield across *HDA20* population.

Figure 9: Selected rootstocks yield compared to non-grafted reference across experiments and scions.

Figure 10: Bivariate plots correlating between variation in genome size and genetic relatedness (a) or sequencing coverage (b).

Figure 11: Alignment of *de novo* genome assemblies of two lines from our core panel. Each point represents a sequence contig.

Figure 12: Examples of potential genomic structural variations between Dulce and TamDew.

Table 1: Properties of 25 *de novo* genome assemblies.

sample	contig_N	contig_N50	genome_size	ONT_coverage2020	genes_liftover (liftoff)
ARJ	289	5,396,572	362,562,842	54	28,208
AY	304	6,854,620	364,934,795	49	28,226
BAHC	248	5,385,344	358,071,711	41	28,227
BDR	235	6,922,960	364,167,905	75	28,235
BEL	382	4,367,020	363,133,380	46	28,198
DOYA	328	5,489,011	365,072,578	63	28,201
DUD	284	6,663,795	360,306,953	64	28,215
DUL	190	7,087,704	365,228,175	64	28,203
ESL	604	2,103,876	356,814,584	56	28,221
INB	268	5,397,639	362,807,472	60	28,185
KRY	393	8,760,176	366,828,714	42	28,214
NA	209	6,964,142	363,899,807	31	28,228
NUN6131	616	1,816,750	363,478,444	39	28,220
NY	204	9,842,676	363,839,612	36	28,217
OHG					
PI161	319	5,650,226	357,638,577	61	28,236
PI164	1,229	905,350	359,844,768	40	28,201
PI414	188	8,741,323	360,708,796	34	28,212
PSR	348	4,637,419	363,084,310	47	28,232
QME	339	4,526,478	358,612,211	31	28,212
SAS	255	7,635,158	357,211,871	53	28,222
TAD	386	3,447,205	367,313,840	43	28,256
TOG	397	3,847,184	359,246,751	40	28,228
TVT	317	4,298,391	364,965,846	73	28,243
VEP	264	5,707,622	364,679,252	64	28,219

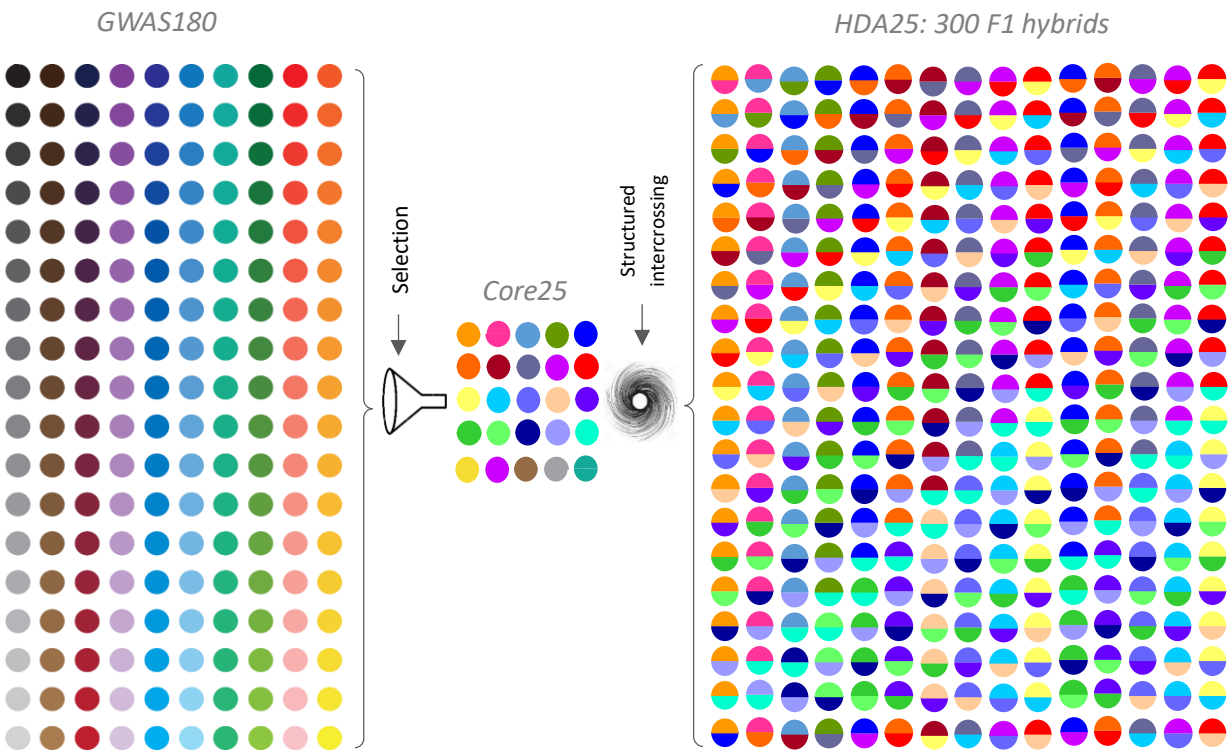**Figure 1**: Our path for development core panel and half-diallele population.

Figure 2: Mode of inheritance of non-grafted **(a)** and rootstocks (grafted, **b**) yield across 45 Half-diallele hybrids and parental lines – Summer 2017.
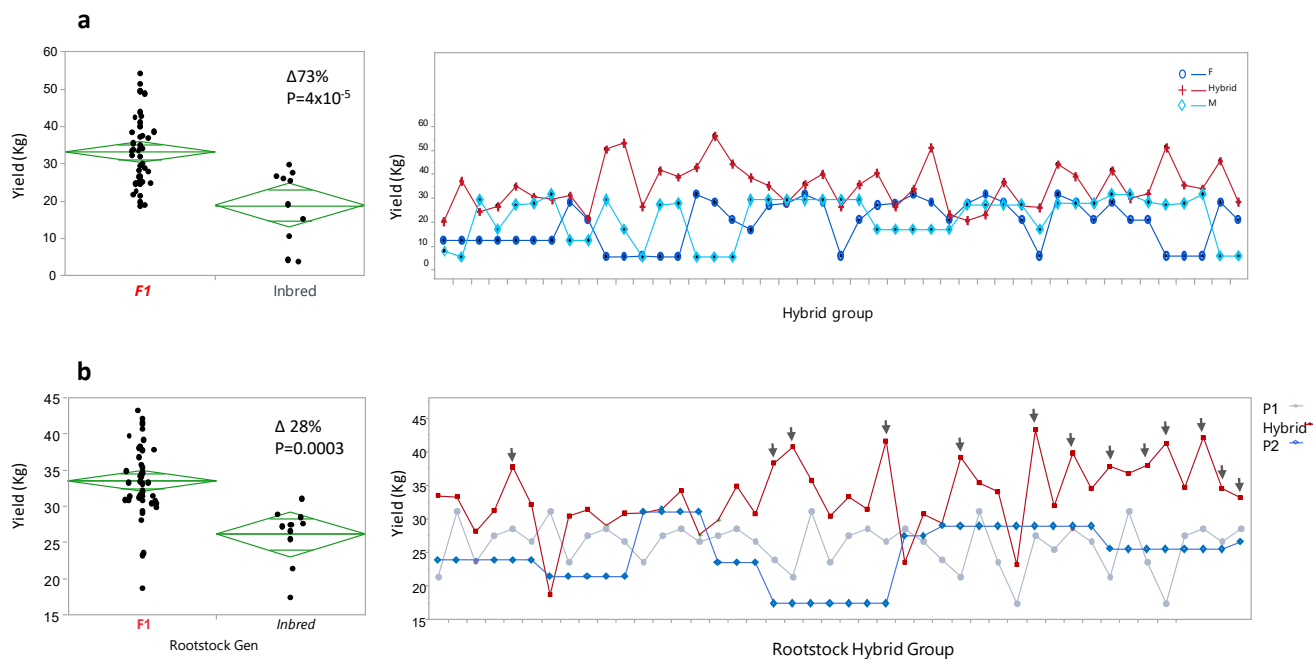

Figure 3: *HDA20* non-irrigated and irrigated yield trials in summer 2018. **a)** Our field at Newe Ya'ar during yield harvest. Melon piles are the yield of plots of five plants. **b)** Yield heatmap projected on the 1,462 field plots (7,310 plants) of the dry and irrigated experiments. **c)** Correlation between dry and irrigated trials. Each dot represents an entry mean in the dry and irrigated fields. Red dots are hybrids and white are inbreds. The common scion, 'Glory', grafted on itself (Gr) and non-grafted (NG) are highlighted.
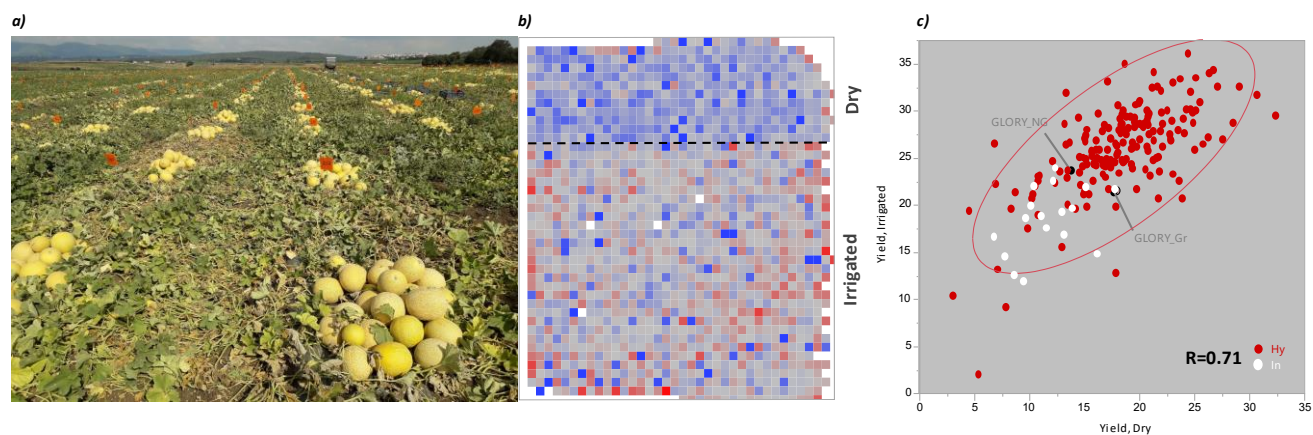

3

Figure 4: *HDA20* grafted rootstock yield trial. Mean comparisons between inbreds and hybrids in the irrigated **(a)** and non-irrigated **(b)** experiments.
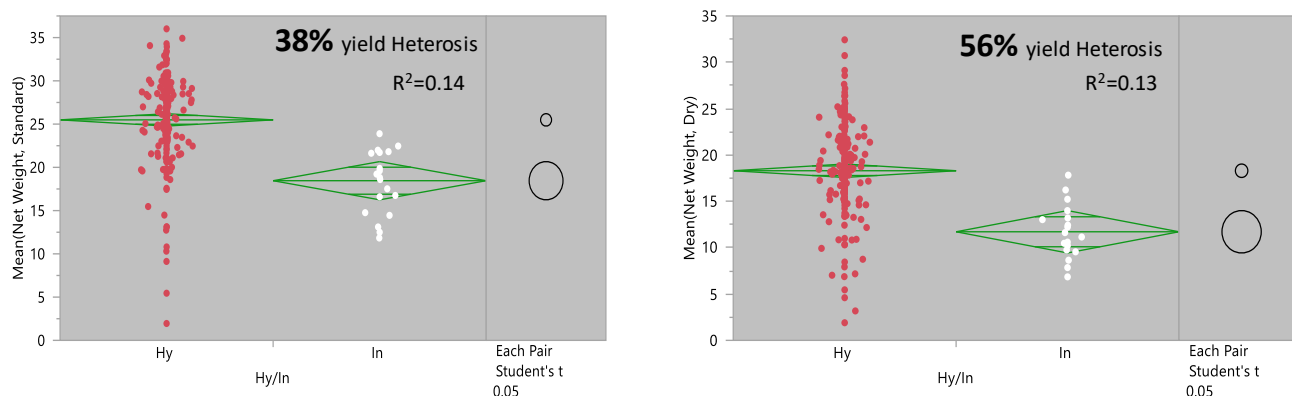

Figure 5: Partition of hybrid yield to parental and heterotic components. **a)** Yield of the 190 *HDA20* hybrids presented by its components: blue bars are the best-parent yield for each hybrid and the orange part represents the deviation of hybrid from best parent (best-parent heterosis; BPH). Hybrids are ordered in an ascending manner by their yield. Negative orange bars reflect hybrids that are lower than their best-parent. 'Glory' grafted on itself is shown with arrow. **b)** BPH effects, presented as Δ% from best parent yield, for each of the hybrids above.
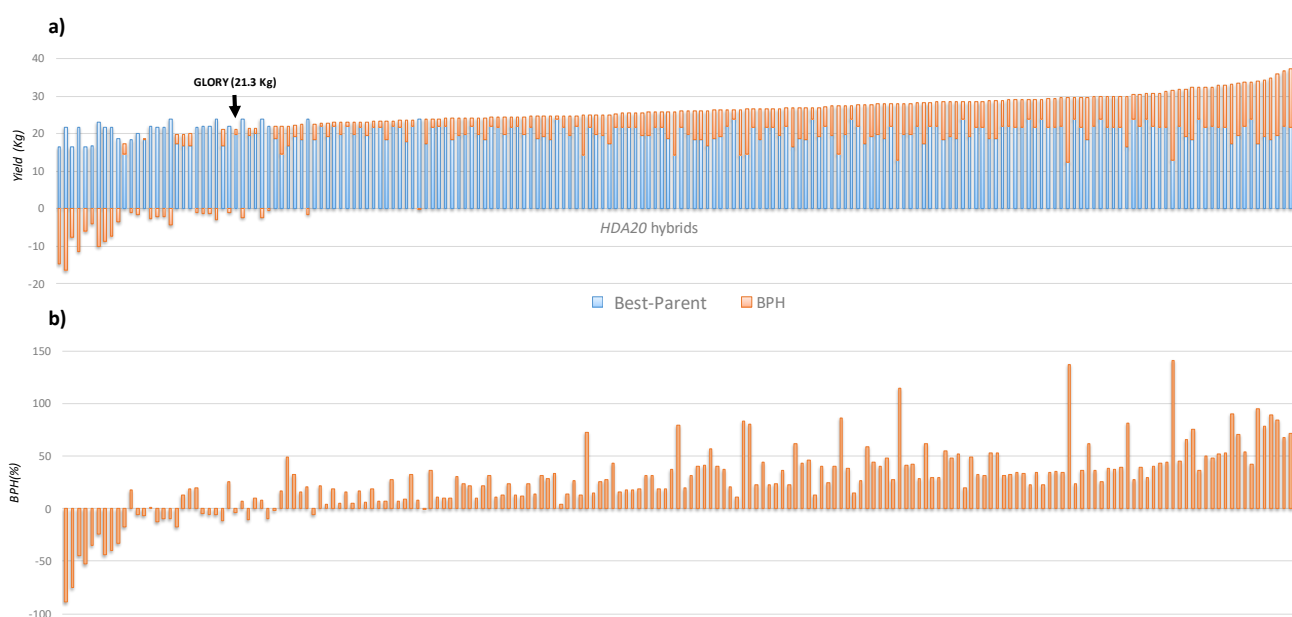

4

Figure 6: Genetic components of hybrid yield in the *HDA20* grafted rootstock experiment. **a**) Comparison between mean yield of all 190 *HDA20* hybrids and their best parents (BP). Each dot represents the average yield of five replicated plots per entry (7 plants/plot). **b**) Bivariate plot of best parent (BP) against F1, across all 190 *HDA20* F1 hybrids. Each point represent the average of 5 plots in the irrigated experiment. X=Y diagonal represent complete dominance (hybrid=best-parent)
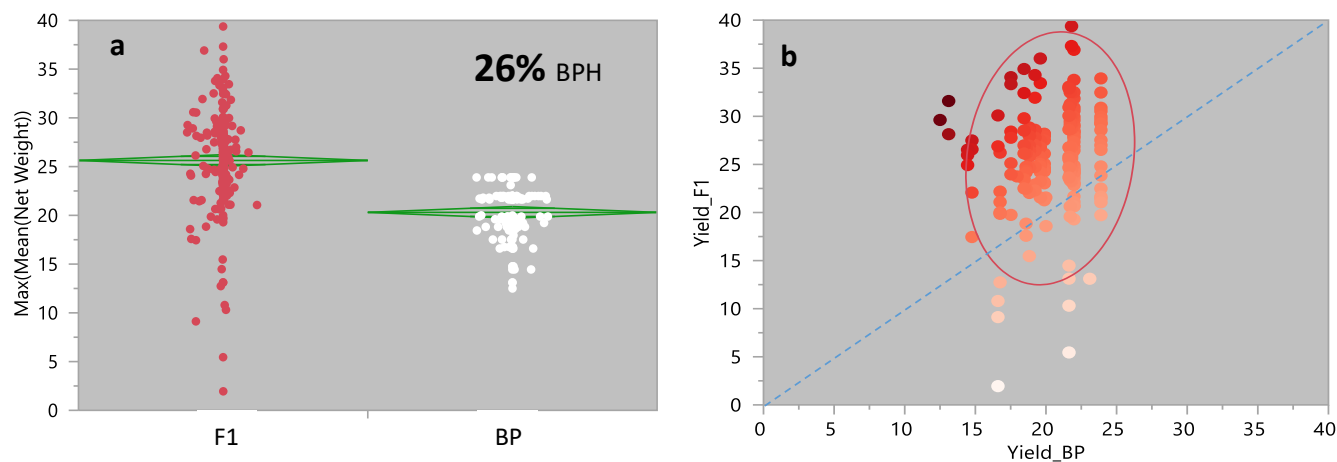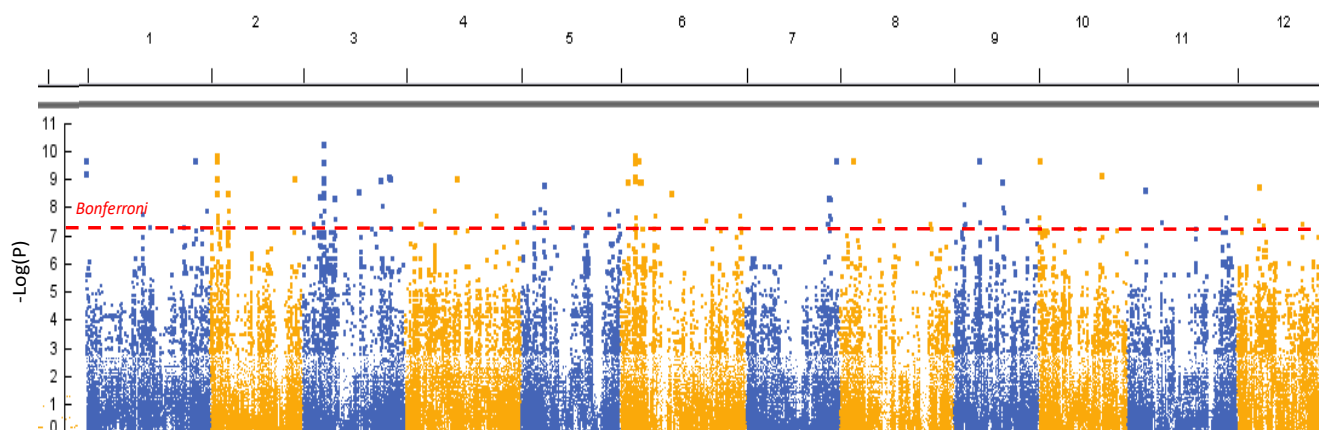

Figure 7: GWAS of root-derived yield across 190 *HDA20* F1 hybrids. Manhattan plot for associations significance of 500,000 genome-wide SNPs. Chromosome numbers are indicated above.

500,000 SNPs, Population-structure corrected model

5

Figure 8: Combined effect of two QTLs associated with root-derived yield across *HDA20* population. The dashed red lines represent the effect of genotypic selection for favorable haplotype at the two QTLs (heterozygote at both loci) on root-derived yield performance.
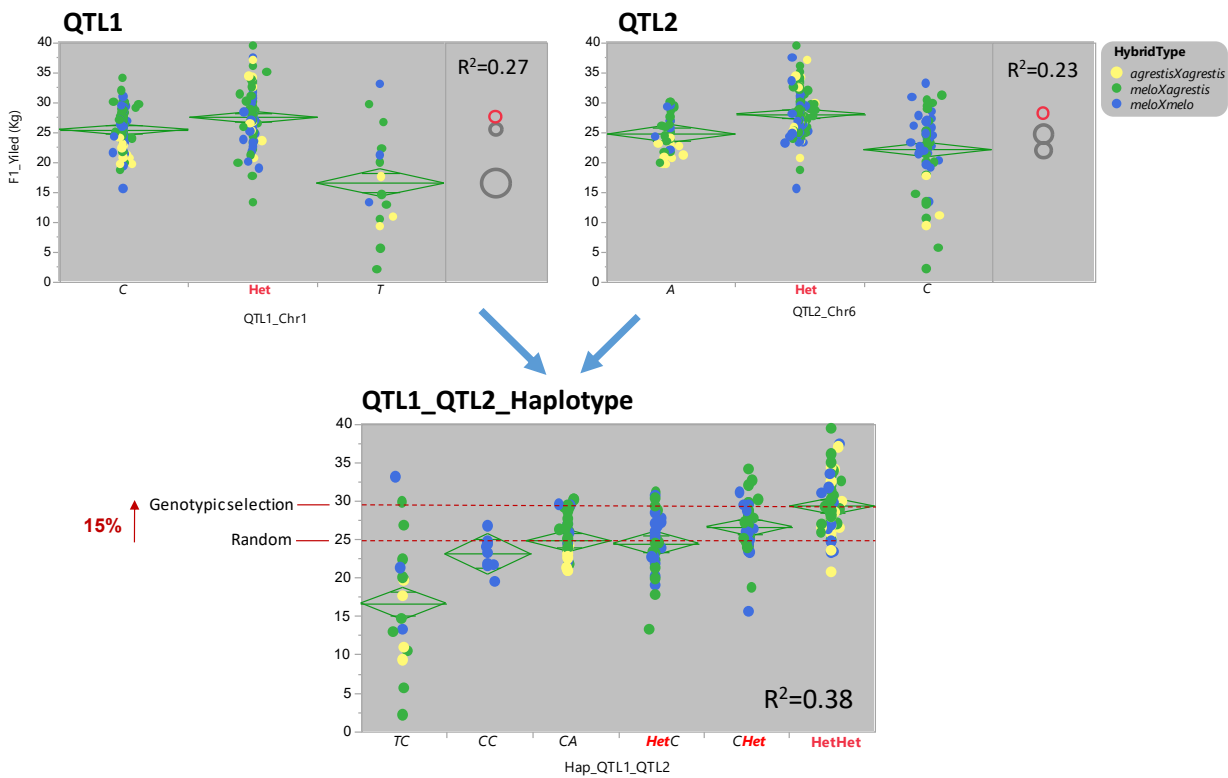

Figure 9: Selected rootstocks yield compared to non-grafted reference across experiments and scions. Values are presented as % difference compared to the non-grafted scion variety (0%), under each environment or treatment.
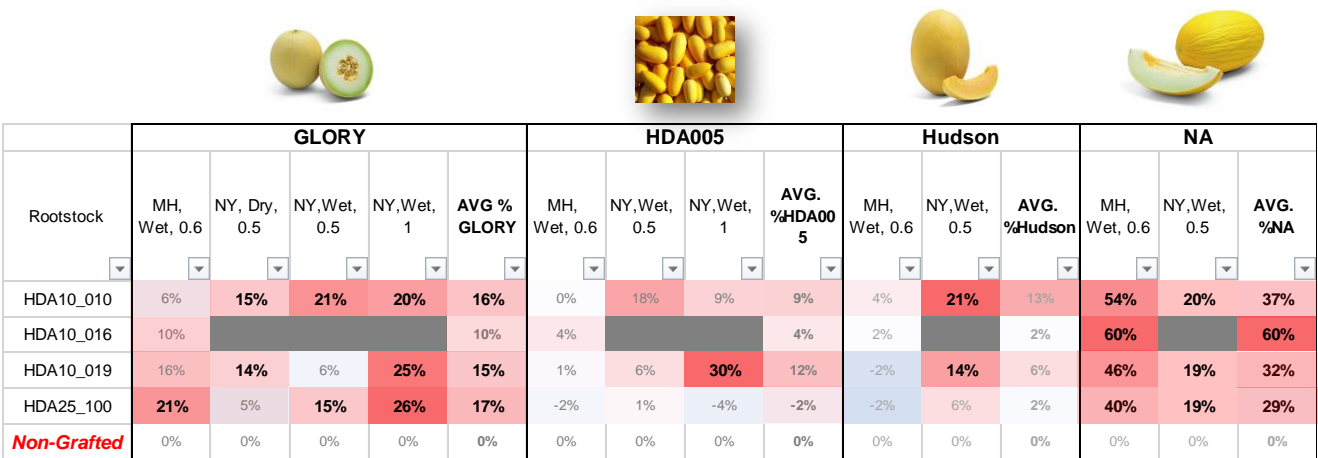

Rootstock	GLORY					HDA005				Hudson			NA		
	MH, Wet, 0.6	NY, Dry, 0.5	NY,Wet, 0.5	NY,Wet, 1	AVG % GLORY	MH, Wet, 0.6	NY,Wet, 0.5	NY,Wet, 1	AVG. %HDA005	MH, Wet, 0.6	NY,Wet, 0.5	AVG. %Hudson	MH, Wet, 0.6	NY,Wet, 0.5	AVG. %NA
HDA10_010	6%	15%	21%	20%	16%	0%	18%	9%	9%	4%	21%	13%	54%	20%	37%
HDA10_016	10%				10%	4%			4%	2%		2%	60%		60%
HDA10_019	16%	14%	6%	25%	15%	1%	6%	30%	12%	-2%	14%	6%	46%	19%	32%
HDA25_100	21%	5%	15%	26%	17%	-2%	1%	-4%	-2%	-2%	6%	2%	40%	19%	29%
Non-Grafted	0%	0%	0%	0%	**0%**	0%	0%	0%	**0%**	0%	0%	**0%**	0%	0%	**0%**

Figure 10: Bivariate plots that show that variation in genome size is associated with genetic relatedness **(a)** and independent of sequencing depth **(b)**.
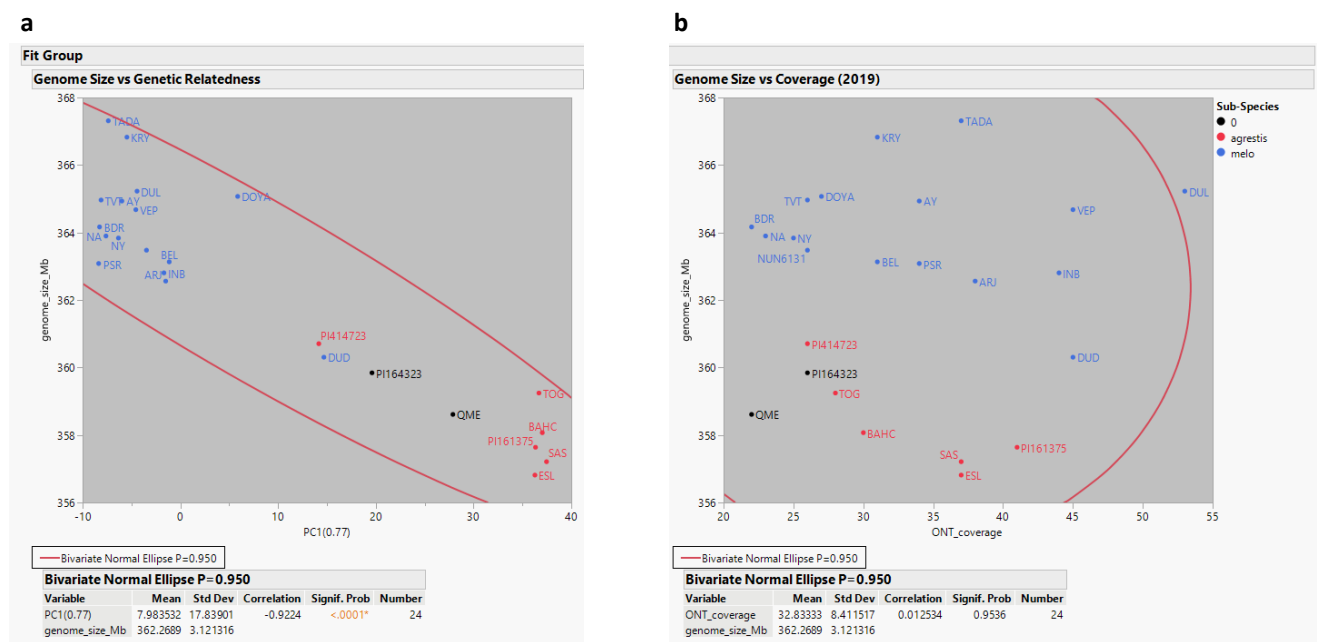

Figure 11: Alignment of *de novo* genome assemblies of two lines from our core panel. Each point represent a sequence contig.
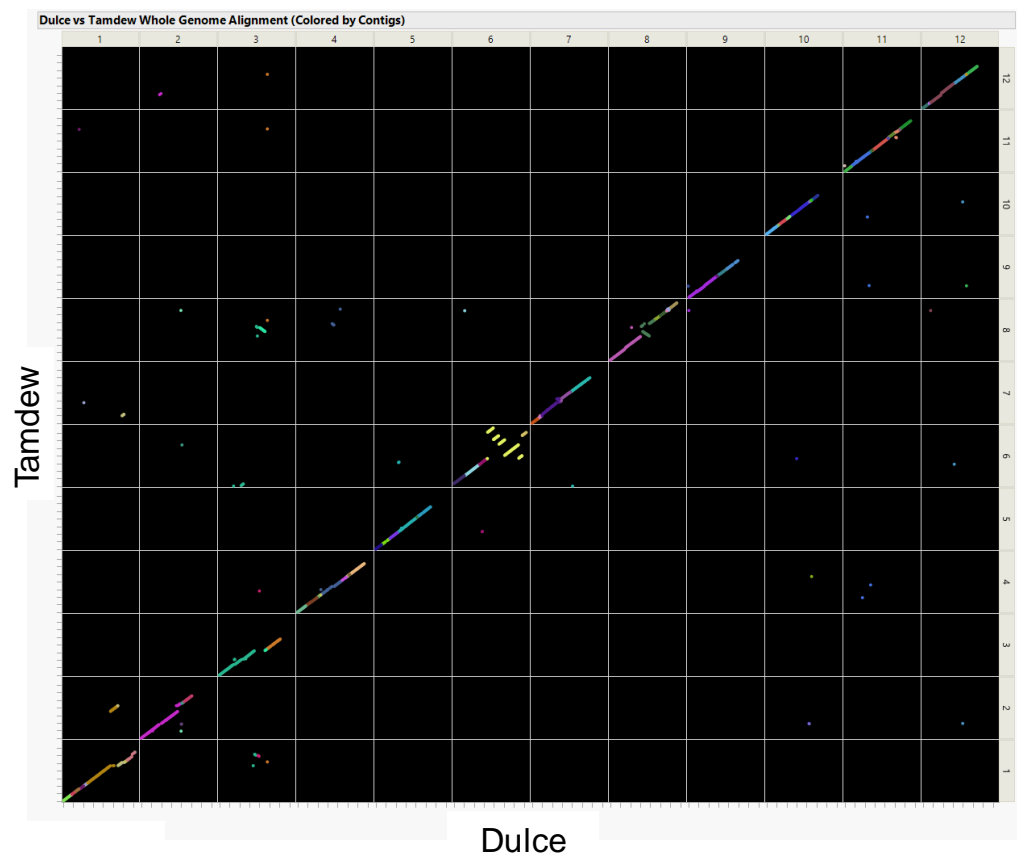

7

Figure 12: examples of potential genomic structural variations between Dulce and TamDew